SPECIAL ARTICLES

THE ORDER OF SCIENTIFIC MERIT

IN selecting groups of one thousand American men of science for statistical study in 1903 and 1909, the workers in each of twelve sciences were arranged in the order of merit for their work by ten of their leading colleagues.[1] The average positions gave the order, and, as there were ten observations of the position of each individual, its probable error could be calculated.

As the writer of this paper is a psychologist and the ultimate object of the work is the study of behavior with a view to advancing scientific research, the psychologists may be used for illustrations of method. William James was placed first in 1903 by the independent judgment of each of the ten observers. The psychologist who stood second had an average position of 3.7 with a probable error of 0.5; the chances are even that his position was between 3.2 and 4.2. The psychologists who stood third, fourth and fifth were assigned, respectively, positions of 4.0, 4.4 and 7.5, with probable errors of 0.5, 0.6 and 1.0. It follows that the relative order of Nos. II, III and IV is not determined definitely, whereas the chances are some 10,000 to 1 that each of these stood below No. I and above No. V. The probable errors increase in size as the work of the men becomes less significant; it is on the average 0.65 places for the first ten of the fifty psychologists and 10.7 places for the last ten. Consequently No. XL on the list would have about one chance in four of falling out of the group of fifty, if the number participating in the arrangement had been very large.

The figures determine not only the validity of the positions, but also the differences in scientific merit among the psychologists, these varying inversely as the probable errors. As men who are about 6 ft. 2 in. tall are likely to differ from each other about ten times as much as men who are about 5 ft. 8 in. tall, so the more distinguished scientific men at the top of the list differ from each other about ten times as much as those toward the bottom, and a unit can be adopted for measuring the differences. This method for converting relative positions into degrees of quantitative differences, which was first used by the writer[2] to measure subjective differences in the intensity of lights has proved to be of wide application.

For a third selection of our thousand leading men of science it seemed desirable, in order to avoid the inbreeding that might occur through selection by a limited group, to obtain a general vote from those competent, and, as before, the new methodological problems have proved to be of interest. The validity of votes appears not to have been considered, yet the problem is wide-reaching and is closely related to the drawing of balls from an urn, which has largely occupied students of the theory of probabilities.

If, for example, the council of the American Psychological Association, which consists of eight members, decides without consultation in favor of a given measure by a vote of 6 to 2, how likely is this to represent the majority opinion of the 432 members? We do not know the distribution of this "population," but if from an urn containing 216 white and 216 black balls, 8 are drawn, the chances are about one in nine that 6 will be white and about one in seven that 6 or more will be white. These may be regarded as approximately the chances that when the membership is about evenly divided a vote of that character will be obtained from the council; and on this basis the desirability of a plebiscite vote may be decided.

In 1916 Mr. Wilson was elected president by the vote in California, which was in his favor by 466,300 to 462,394, giving him a majority of 3,906. It might be supposed that this small majority would readily have been reversed by an indefinitely large electorate, but if the population were equally divided this would probably occur only once if an election were held every day for a hundred years.

If the members of a jury reached their deci-

[1] SCIENCE, November 23, November 30 and December 6, 1906, "American Men of Science," The Science Press, 1910.

[2] *Philosophische Studien,* 1902. See also "Studies by the Method of Relative Position," H. L. Hollingworth, in "The Psychological Researches of James McKeen Cattell," a review by some of his pupils, on the occasion of the twenty-fifth anniversary of his professorship. New York, 1914.

sion without consultation and stood eleven to one for conviction, many would conclude that the chances are eleven to one that the defendant is guilty. As a matter of fact if the total population is divided in the same ratio and the legal fiction is followed that a man is proved guilty or innocent only by unanimous vote of twelve peers, the chance of obtaining a jury which without consultation will be unanimous for conviction are about one in three, of obtaining a jury unanimous for acquittal, only one in many billions. The practise of the courts must ultimately adjust itself to such conditions, and learn whether the unanimous vote of three or five jurymen without consultation is more or less valid than the unanimous vote of twelve after consultation. It must be decided on what probabilities a man shall be convicted and to what extent the chances of innocence shall be considered in imposing sentence.

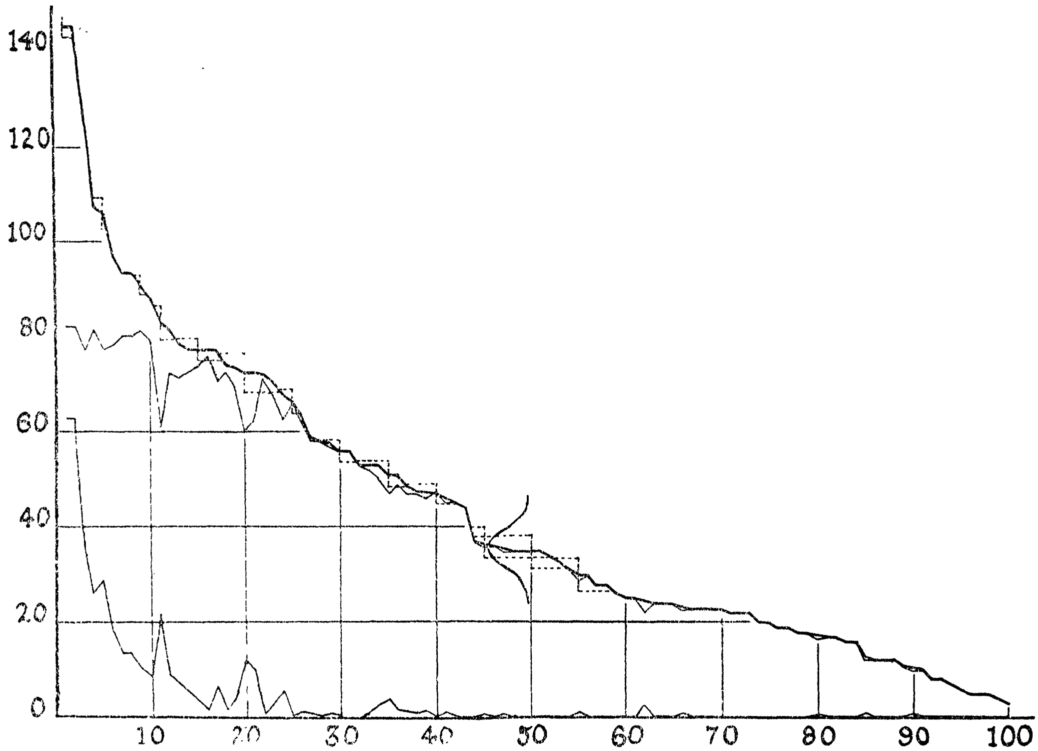
When our rulers are selected and their legislation is determined by votes, it seems strange that knowledge concerning the variable and constant errors is so completely lacking. A decision of the supreme court, which may involve the welfare of hundreds of thousands of children or taxation amounting to billions of dollars, is equally binding whether the vote is unanimous or by a majority of one. It would apparently be as reasonable to require a three fourths vote of the supreme court to invalidate a law passed by both houses of the congress and signed by the president, as to require a unanimous vote of a jury to award petty damages.

The method of voting used to select scientific men might have useful applications in industry. For example, if a bank employs 100 clerks, some of whom will be promoted from time to time, all of them might be asked to check the fifty per cent. and the five per cent. most deserving of promotion. From the records an order of merit for promotion would be obtained, together with the relative value of the men to the bank and the salaries deserved. The data would also throw light on those voting, for the value of the judgment of each is measured by its departure from the average; if any were prejudiced or unfair in their treatment of friends or rivals this would be discovered. The combined judgment of associates is probably a more valid method of selection than the choice of a superior and would presumably lead to better service.

In like manner a group of factory workers or of laboring men might select a boss or leader by vote. If the employer would agree to take one of the five men receiving the most votes, the employees would understand and probably approve the method. Not only would a good selection free from favoritism be made, but the men would share in the control of their work and would be more loyal and more efficient.

The writer has proposed a compromise between the competitive and the semi-communistic systems for payment of the salaries of university professors, according to which, say, five super-professorships with relatively large salaries and large freedom should be established. The difficulty under our present method is that the appointments would be by favor of the administration. If, however, a vote of the teachers in the university, and perhaps of the students, were taken on the basis of desert for research, teaching and service, a method of selection would be used probably more accurate than the choice of the president and at the same time more conducive to cooperation and goodwill.

In the selection by votes of one thousand scientific men the same number for each of the twelve sciences was retained as in the two previous studies, this being nearly proportional to the total number of workers in each science. In the first edition of "American Men of Science" there were listed about 4,000 scientific men, in the second edition about 5,500, and in the present edition, published some fifteen years after the first, the number is about 9,600. The present writer was in 1888 the only professor of psychology in the world. The number of members of the American Psychological Association, which has a professional qualification, increased from 127 in 1903, when the first selection was made, to 432 in 1920. The average number of doctorates conferred in psychology during the five years prior to 1903 was 12.6; it was 40 in 1920. Competition for inclusion among our fifty leading psycholo-



THE DISTRIBUTION OF VOTES FOR ONE HUNDRED PSYCHOLOGISTS

gists or among our thousand leading men of science is thus much more severe at present than was the case when the selections were first made. There are now about 500 working psychologists and about 10,000 scientific men in the United States; the present selection consists approximately of those who form in each science the upper tenth in merit of their work.

In the case of the psychologists, here used as an example, those in the two earlier groups of fifty (except five no longer occupied with psychology or no longer residents of the United States), numbering 48, were asked to send the names of 10 or a smaller number of others whose work warranted their inclusion in such a group. Then those who received two or more nominations were in like manner asked to propose 10 or fewer names. The 52 who received the most votes were added to the original 48 to form a group of 100. These names were placed in alphabetical order on a list, which was sent to them all, with the request to "check (√) about 50 (namely, about one half) of the names to indicate those who have done the best work in psychology, placing a double check (√√) before about five of those whose work has been the most important." Eighty-three (besides one who replied too late) of the 100 returned the blank and, as each was asked not to consider himself, there were 82 votes. When more than five double checks were assigned they were weighted inversely as the number. The result of the votes can be conveniently shown by curves such as are here reproduced. The middle curve gives the distribution of the 82 votes for the 100 psychologists, the vertical ordinates representing the number of votes, while the individuals are ranged serially along the horizontal axis. The bottom curve shows the number of double checks for each individual, namely, the opinion that he is one of the five psychologists whose work is the most important. The top curve represents the sum of two votes, and the individuals are arranged in order in accordance with this vote.[3]

[3] Similar results have been obtained in each of the twelve sciences, the number of scientific men

As indicated on the curve and shown on the tables (which it seems unnecessary in this place to print) two and only two psychologists were included among the fifty leading psychologists by unanimous vote of their eighty-two colleagues. The psychologist at the bottom of the fifty received 35 votes; three were given for the psychologist who stands last in the hundred. Sixty-two of the 82 voting include among the five whose work has been most important the two psychologists who received a unanimous vote for inclusion among the fifty and 20 do not. Such differences in judgment are legitimate and significant. Thus the psychologist placed eleventh is held by 22 of his colleagues to belong to the first five and by 21 not to belong to the group of 50, and similar conditions obtained for the one placed twentieth. The names could be guessed by one familiar with the situation. They are men of distinction whose more important work is by some judged not to fall within the field of psychology.

The attitude of those voting is of scientific significance, for it measures the validity of judgments. If we assume the average judgment of the 82 psychologists to be nearly correct, the departure from this average measures the competence of the individuals to form such judgments. There do not appear to be group differences dependent on distinction or age, but individuals, as shown in the previous study, differ in the ratio of about two to one. There will be an extraordinary change in our attitude toward political, social and business problems when we learn to look upon our observations, recollections, beliefs and judgments objectively, measuring the probability of their correctness and assigning probable errors to them.

The top curve represents the sum of the two votes and the order there given is the one used. A different weighting of the two votes would not considerably affect the order. The vote for the five leading psychologists in the main discriminates only the positions of the men in the upper quartile. In place of the double vote for the five and the fifty per cent. of the 100 psychologists whose work has the most merit, a satisfactory distribution might be obtained by a vote for 25, or one fourth of the whole number. In view of the constant use of votes for elections and decisions, the problems involved deserve more complete investigation.

A probable error can be found for the positions of the individuals by a method that was apparently first used by the present writer. When eighty of those voting are divided into ten groups of eight each, we have the separate votes of each of these groups and from their variation the probable error of the average vote can be calculated. Thus the psychologist No. L, in the ten groups of eight votes each received, respectively, 4, 4, 2, 1, 4, 7, 4, 3, 2, and 2 votes. The probable error is 0.363, and for the group of eighty votes it is 3.63. The position on the curve assigned by each of the small groups can also be found and a probable error calculated from these ten positions.

The probable errors based on 80 votes (as a rule for each fifth individual) are indicated by the broken vertical line on the curve and when referred to the order of merit by the broken horizontal lines. The probable errors of the votes of the five psychologists last in the fifty are, respectively, 3.4, 2.8, 2.8, 3.1 and 3.6, an average of 3.14. The curve from No. X to No. C is nearly a straight line, the vote decreasing from 79 for No. X to 3 for No. C. Consequently the probable error of the vote when referred to the order is increased by

ranging from 175 chemists to 20 anthropologists. The returns, however, were most complete for psychology, the writer being personally acquainted with nearly all psychologists and a second request having been sent to those who did not reply to the first. There were in all some 130,000 votes to be collected, counted and tabulated. For the treatment of this material I am mainly indebted to my daughter, Miss Psyche Cattell. Dr. Dean R. Brimhall and Dr. Alexander Weinstein have also assisted in the revision of the material and in the computations.

I am under very great obligations to Professor Raymond L. Pearl, of the Johns Hopkins University, and to Professor H. L. Rietz, of the State University of Iowa, for their kindness in reading the manuscript of the paper and for the useful suggestions that they have made. This acknowledgment should not, of course, be construed as involving responsibility on their part.

about one sixth. The probable error of position at the bottom of the group of fifty is 3.6; there are thus only three or four for whom there would be as much as one chance in four of being dropped from the list if the arrangement were made by an indefinitely large electorate of the same character.

In the first study of the psychologists, the probable error at the lower end of the fifty was 10.7, that is, there were 10 or 11 for whom there was one chance in four that they should not be included on the list. There were then ten arrangements in order of merit; now there are eighty votes. The probable error decreases as the square root of the number of observations, and the probable errors in the two cases, other things being equal, should be about as 3 : 1, which is in fact almost exactly the case. The probable error of a single vote for the psychologists low in the group of fifty at the present time is thus the same as by order of merit in the group of 1903.

In the first study, however, we were concerned with the upper quartile, and we are now concerned with the upper decile of the group of American psychologists. If we assume distributions in accordance with the curve of error, the men who now stand at the bottom of the fifty in the present selection will be as able as those who stood about twenty in the first arrangement. An examination of the relative positions of the individuals who are in each of the two arrangements indicates that this tends to be the case. The probable error at the bottom of the fifty selected by votes should be in the neighborhood of those who stood about twenty in the first arrangements. This would make a single vote one half as valid (the ratio of the probable errors in the first arrangement for those near the bottom of the fourth hundred and in the tenth hundred of the thousand being 64 : 125) as a single judgment of order of merit. The figures given should, however, be regarded as indications of method rather than as exact determinations, for they are subject to various errors.

The average position of the survivors of the first group of psychologists in the arrangement of 1903 was (after deaths have been eliminated) 4; it is now for the same individuals 14.8. Those in the four following groups of ten have dropped, respectively, from 11.5 to 18.1; 20 to 27.4; 29.5 to 59.2, and 39.5 to 64.9. This drop in position is on the average less than would be expected if the 10,000 scientific men of to-day are as able and have as good opportunity as the 4,000 scientific men of 1903. The inference is that as the total number increases the proportion of men of distinction decreases. This may be due to the fact that men of special ability find their level apart from the size of the group or because the scientific career attracts less able men or gives them less opportunity than formerly. Both factors are probably present; it is apparent that the situation deserves further investigation. In the previous study it was shown that in the increased competition of a five-year period, those between 40 and 44 years of age remained on the average about stationary; those below 40 gained; those above 44 lost, the loss being in direct proportion to the age.

As the work of the men becomes more important, the differences between the individuals as measured by the probable error of position become greater, the distribution corresponding in a general way to the upper end of the curve of error. In the case of the votes there are complications, for the votes for different men do not have the same weight. The ten who receive the most votes receive nearly all the votes, and in the cases of the few who do not vote for them poor judgment or an error in checking is indicated. Some votes mean that a psychologist stands first or near the top of the list, whereas others mean that he is barely included. Consequently the order and the probable errors in the case of such a vote for one half of the group do not have great validity for the upper part of the distribution. The order is obtained in a satisfactory manner by the double vote, but this introduces further complications in the probable error. In all cases of votes, we have asymmetrical distributions and skew curves. The quantitative relations should be worked out in the first instance for less complicated material than that with which we are here concerned.

In nearly all cases in which probable errors have been applied to psychological data, the

determinations are more exact than common sense would presuppose. Thus the writer found[4] that in grading traits of character by ten individuals on a scale of 100, a position was assigned with probable errors varying from 4.6 for physical health and cheerfulness to 3 for originality and efficiency. All other traits, such as energy, courage, judgment and integrity, were assigned positions with intermediate probable errors, the average being 4, which is nearly the same as the probable error of position as determined by 80 votes of the psychologists near the middle of the 100 in order of merit.

The comparatively small probable errors appear to be due to the fact that there are constant errors which affect the whole group. The psychologists who vote are subject to the same kind of influences, not making in fact independent judgments, but being influenced as a group by the knowledge of what others think and by all sorts of conditions, conventions and restrictions. If a similar vote were taken ten years hence the work of the same psychologists would be viewed from new standpoints and the positions would change to a much greater degree than the probable errors warrant. "Constant" errors are in fact more inconstant and variable than "variable" errors.

In the case of a vote (as in any series of measurements) there are two factors entering into the probable error, one dependent on the quantitative conditions prescribed in advance, the other on the behavior of the individuals. The former may be called the deductive probable error and when the latter is determined by experiment and added to it the whole is the inductive or actual probable error. Thus, if from an indefinitely large number of balls equally distributed between black and white, some are drawn, the most frequent distribution will be an equal number of black and white, but the average departure from equality will increase as the square root of the number drawn and the ratio of departure from equality will decrease as the square root of the number.

If large numbers of white and black balls are distributed in the ratio of 33 white to 47 black, and we draw 80 balls, the most probable number of white balls will be 33. The standard deviation from 33 in a large number of draws will be 4.40, and the quartile deviation or probable error will be 2.97; that is, in one case out of four there will be more than three white balls. The psychologist at the bottom of the fifty received 33 votes out of a possible 80. If an indefinitely large number of psychologists were distributed in this ratio the deductive probable error or error of sampling would be 2.97. The actual probable error, namely, 3.63, is composed of (the square root of the sum of the squares of the two) this deductive probable error and an error or deviation due to the groupings of the psychologists into different "species" with different points of view. The psychologist who stood XIL had a probable error of 3.1. The deductive probable errors are approximately the same for the two individuals, but it is more difficult to form a judgment regarding No. L than regarding No. XIL.[5]

The situation may be illustrated by an instance of general importance. Death rates, birth rates and marriage rates are continually used, but always without probable errors. Thus, for example, the Bureau of the Census issues weekly a bulletin that contains the death rates of the leading cities of the United States, but the figures have no meaning because one does not know whether the different rates are due to chance fluctuations with a limited population or to causes such as a large proportion of infants or an epidemic of influenza.[6] If the

[4] Address of the president of the American Society of Naturalists, SCIENCE, April 10, 1903.

[5] In these cases the actual and the deductive probable errors have probable errors of the order of magnitude of the differences between them, and these differences have only moderate validity. The writer has purposely ''not minded his p's and his q's,'' for it seems that equations are not becoming to one who is not a mathematician.

[6] In the last report received (for the week ending September 2, 1922) the death rate of New Haven is given as 5.8 and of Houston as 13.9. In the same week a year ago the death rate of New Haven was 10 and of Houston 7.6. Without probable errors these figures give no useful information in regard to the conditions in the two cities.

average death rate is 12 per thousand, in a city of 100,000 population there will be about 23 deaths in a week. If black and white balls in indefinitely large numbers are distributed in the ratio of 23 black to 99,977 white and 100,000 are drawn, the most probable number of black balls is 23, but one time in four there will be more than 27. Thus the recorded death rate for a week for a city of 100,000 will normally fluctuate. If it is on the average 12, it will in half the weeks be approximately either as large as 15 or as small as 9.

If the death rate exceeds 15 in two consecutive weeks then the chances are fifteen out of sixteen that it is due to some cause such as an epidemic. The conditions are obviously of practical importance for physicians and health officers. The situation for death rates is nicely illustrated by the illustration that has been used of the distribution of black and white balls in an urn. If the population of the country were 100,000,000 and the death rate were 12 (as it should be, but is not), then 1,200,000 people would die during a year. Among 100,000,000 black and white balls there are 1,200,000 black. But if we draw 100,000 (*i. e.,* take a town of that population) there will be a chance fluctuation as described above. It is also the case that the balls are not completely mixed, there being more black balls in some part of the urn than others. In some places we shall draw a larger proportion of black balls. When there is a negro population or a tenement house population or a large population of very young or very old people, there are relatively more black balls. There are temporarily more black balls in one place when there is an epidemic or the like. In that case we have the analogy of the black balls attracting one another.

This paper has been written to explain the methods used to select the thousand leading American men of science by votes. The psychologists have been taken as an example; if space and time permitted tables and curves might be given for the other sciences and a study of the data might yield results of interest. Such treatment must, however, be postponed or left to others. The object of the present paper will be accomplished if it makes clear that the scientific men have been selected and placed in the order of merit for their work by valid objective methods and that the methods used have wide application. In a subsequent paper the distribution of the scientific men will be considered with special reference to the changes that have occurred in the course of ten years.

J. McKEEN CATTELL
THE PSYCHOLOGICAL CORPORATION,
AUGUST 1, 1922

THE AMERICAN CHEMICAL SOCIETY

(Continued)

DIVISION OF LEATHER CHEMISTRY

John Arthur Wilson, *chairman*
Arthur W. Thomas, *secretary*

The mechanism of unhairing: JOHN ARTHUR WILSON and GUIDO DAUB. A series of detailed studies was made of the mechanism of the unhairing of skins by means of the sweating process, lime liquors and caustic sulfide liquors. Sections of skin were examined under the microscope at different stages. In liming and in sweating, the first action on the skin itself is the hydrolysis of the epithelial cells of the Malpighian layer of the epidermis, once the cells are destroyed, the remainder of the epidermis, the hair and the sebaceous and sudoriferous glands are completely separated from the derma and can then be removed mechanically. In the sulfide method, the alkali destroys the corneous layer of the epidermis and the skin appears to be freed from epidermal matters on its surface long before the alkali has penetrated to the depth of the hair bulbs. Where this method has been employed, the hair bulbs are usually found intact in the finished leather. The paper is illustrated with photomicrographs.

Pancreatin as an unhairing agent: JOHN ARTHUR WILSON and ALBERT F. GALLUN, JR. When calf skin is swollen in dilute caustic soda, neutralized with sodium bicarbonate, and then put into a suitable solution of pancreatin at 25° C. exposed to air, the hair is completely loosened in 24 hours, but the action is not due to the enzyme, since it is checked by covering the solutions with a layer of toluene. At 40° a solution of pancreatin fails to cause a loosening of the hair of fresh skin because the corneous layer of the epidermis is impermeable to the enzyme and